



Optimization of data acquisition operation in optical tomography based on estimation theory

MAHSHAD JAVIDAN,^{1,2} HADI ESFANDI,^{1,2} AND RAMIN PASHAIE^{1,*} 

¹Electrical Engineering and Computer Science Department, Florida Atlantic University, Boca Raton, FL 33432, USA

²Authors contributed equally

*rpashaie@fau.edu

Abstract: The data acquisition process is occasionally the most time consuming and costly operation in tomography. Currently, raster scanning is still the common practice in making sequential measurements in most tomography scanners. Raster scanning is known to be slow and such scanners usually cannot catch up with the speed of changes when imaging dynamically evolving objects. In this research, we studied the possibility of using estimation theory and our prior knowledge about the sample under test to reduce the number of measurements required to achieve a given image quality. This systematic approach for optimization of the data acquisition process also provides a vision toward improving the geometry of the scanner and reducing the effect of noise, including the common state-dependent noise of detectors. The theory is developed in the article and simulations are provided to better display discussed concepts.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Tomography scanners are used in a wide range of non-destructive testing applications. Despite some differences, most scanners are made based on the same architecture in which an object is placed between an array of sources and detectors, Fig. 1. In each test, one or a combination of sources are powered to radiate some form of energy, e.g., optical, electromagnetic, acoustic, thermal, etc., into the medium under test. The radiated energy interacts with the medium as it propagates from sources to detectors, where the intensity and sometimes phase of the field are measured. Scanners collect a set of these measurements and solve an inverse problem to reconstruct three-dimensional images that reveal information about the structure or certain properties of the medium.

The most common scanning protocol in today's optical tomography is raster scanning. In raster scanning, for each measurement, only one source is illuminating power while all detectors are recording. The scanner makes all measurements before using the acquired data to reconstruct the image. Raster scanning is time-consuming and such scanners are occasionally not fast enough to catch up with the speed of changes when imaging dynamically evolving objects [1]. To expedite the process of data acquisition, we need to reduce the number of measurements but make each measurement as informative as possible. By choosing appropriate geometry for the scanner, including the locations of sources/detectors, and engineering the illumination pattern in each measurement, we can improve the rate of data acquisition and reconstruct images with satisfactory quality from a smaller number of measurements. This will also reduce the cost of operation.

Previous work in the field can be divided to two main categories. The first category includes algorithms that are designed based on concepts of signal processing which assumes system identification can be expedited if we measure the response of an unknown system to complex inputs instead of measuring a set of impulse responses such as raster scanning. Different illumination patterns were tested for this purpose, including checkerboard [2], sinusoidal [3], wavelet [4], or diffractive optics patterns [5]. Based on published literature, these methods provide

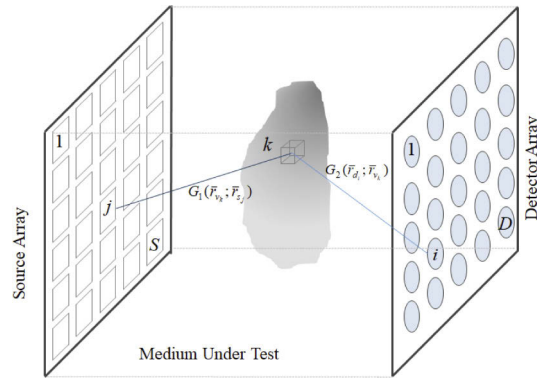


Fig. 1. Schematic of a typical scanner. The medium under test is placed between an array of S sources and D detectors and the medium is divided to V number of voxels. The energy from source j at \bar{r}_{s_j} propagates to all voxels including voxel k at \bar{r}_{v_k} . The energy is scattered by the medium and then detected by all detectors including detector i at \bar{r}_{d_i} .

only marginal improvements. In a separate approach, different scanner geometry or illumination patterns were tested to enhance the information content of the weight matrix which appears in the formulation of the problem as discussed later [1,6–9]. For instance, in one study, illumination patterns were optimized by improving the conditioning of the Fisher Matrix to maximize the rank of the weight matrix [10]. For sensor position optimization, a two-step algorithm was proposed in [11]. In this algorithm, first, the deviation between theoretical estimation and the result of an actual measurement is minimized for a given sensors' position, and the image is reconstructed. Then, a second cost function, which describes the quality of reconstructed image, is optimized to find the optimal sensors' position for the next round of measurements. Similar approach was also used to extract optimal illumination patterns for diffuse optical tomography [12].

This second category of algorithms is proven to be more effective; however, what is missing in previous algorithms is the main asset we have for optimization of data acquisition routines which is our prior information about the object. In fact, many research groups have already utilized prior information to alleviate the ill-posedness of the inverse problem in optical tomography by means of regularization techniques [13–16]. Since Kalman Filter is a powerful tool for employing prior information to solve inverse problems, we came to the idea of viewing this problem as an estimation theory problem. However, in this study, we intend to take advantage of the prior information not only to attenuate ill-posedness of the problem, but also for the optimization of data acquisition process which leads to improved image quality as well. Prior information can be acquired from various sources such as our knowledge about the sample under test, results of initial measurements, data from other imaging modalities, or previously acquired images in our database. Consider the example shown in Fig. 2. In the study of the brain circuitry, we occasionally need to evaluate the success of the gene delivery process. For this purpose, we co-express some form of fluorescence protein to label genetically modified cells. After a short period following the injection of virus, which acts as the vehicle for gene delivery, target brain cells produce the fluorescence protein. The concentration of the fluorescence protein in the area is used to quantitatively determine the success rate of gene delivery. Since the coordination, amount, and timing of the injection are all known variables, we have a good estimation of the spread and volume of the transfected area before making any measurement.

In this article, we consider the discretized sample as a state vector and the covariance matrix quantifies our uncertainty about the concentration and the spatial spread of the target in the sample. We then introduce an optimization method in which the scanner identifies projections

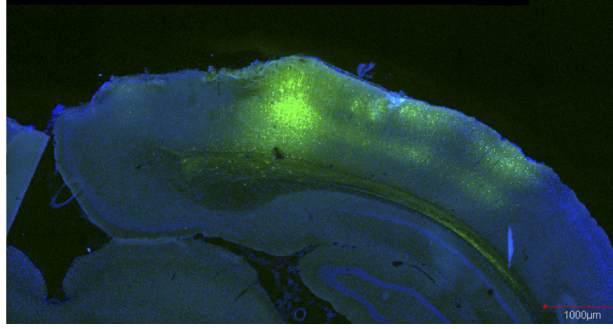


Fig. 2. Expression of green fluorescence protein (GFP) following the procedure for viral gene delivery. Controlled amount of the solution that contains the virus was injected in the cortical area about 500 micron inside the tissue. The solution diffused into the tissue and the virus delivered the genetic construct to the cells which led to the expression of GFP in excitatory neurons. The volume and position of injection can provide a good estimation of the distribution of fluorescence proteins after a given period of time post injection. This information can be used to find the initial estimation prior to any fluorescence tomography test.

of the state vector for which we have maximum uncertainty. Next, the system synthesizes illumination patterns to specifically make measurements along those projections. This statistical approach incorporates the effect of measurement noise and other sources of uncertainties (such as uncertainties in model predictions or state-dependent noise) that influence the performance of the scanner.

2. Mathematical framework

In this article, we study image reconstruction as a state estimation problem with two main ingredients; the first is the forward problem, defining how observations are related to estimation variables, and the second is the mathematical layout which explicitly explains how estimations are updated based on the forward model prediction and partial observations.

2.1. Forward model

The architecture of a typical scanner is shown in Fig. 1. Most tomography systems have well-defined geometries, e.g., a cylinder or sphere. Nonetheless, flexible scanners are also made to conform to the arbitrary shape of any object under test. In all forms of tomography, we need to radiate energy into the medium to remotely explore its structure. Wave propagation in physics is modeled by Helmholtz partial differential equation, which in the scalar case, takes the following form:

$$\nabla^2 \phi(\vec{r}) + \kappa^2 \phi(\vec{r}) = -\eta(\vec{r})\phi(\vec{r}). \quad (1)$$

In this equation, \vec{r} is the position vector, ϕ represents the field intensity, and κ is the wave number which incorporates the effect of the medium on wave propagation. In general, κ is a complex number. When κ is purely imaginary, wave propagation transforms to diffusion. The function $\eta(\vec{r})$ is the scattering potential and finding this function is the main objective of every tomography problem. This equation is usually solved by decomposing the field intensity to the incident, $\phi_i(\vec{r})$, and scattered fields, $\phi_s(\vec{r})$, so that: $\phi(\vec{r}) = \phi_i(\vec{r}) + \phi_s(\vec{r})$.

$$\begin{aligned} \nabla^2 \phi_i(\vec{r}) + \kappa^2 \phi_i(\vec{r}) &= 0, \\ \nabla^2 \phi_s(\vec{r}) + \kappa^2 \phi_s(\vec{r}) &= -\eta(\vec{r})\phi(\vec{r}). \end{aligned} \quad (2)$$

The first equation formulates the intrinsic response of the system in the form of an eigenvalue problem. Information about the medium is embedded in the second equation. If we transform the second equation to its dual integral form, we can compute the intensity of the scattered field at the location of the i th detector, \bar{r}_{d_i} :

$$\phi_s(\bar{r}_{d_i}) = \int d^3\bar{r} G_2(\bar{r}_{d_i}; \bar{r}) \phi(\bar{r}) \eta(\bar{r}), \quad \forall i \in \{1, \dots, D\}. \quad (3)$$

In this equation, the integral is taken over the entire volume of the medium. The function, $G_2(\bar{r}_{d_i}; \bar{r})$ is the Green's function which models the propagation of the field, after interacting with the medium, from the position \bar{r} in the medium, to the position of the i th detector. Within the first Born approximation regime, the scattered field can be computed from:

$$\phi_s(\bar{r}_{d_i}) \simeq \int d^3\bar{r} G_2(\bar{r}_{d_i}; \bar{r}) \phi_i(\bar{r}) \eta(\bar{r}). \quad (4)$$

If we assume that all sources are discrete point sources, then:

$$\phi_i(\bar{r}) = \sum_{j=1}^S I_j G_1(\bar{r}; \bar{r}_{s_j}). \quad (5)$$

The variable I_j is the intensity of the field, illuminated by the j th source. In practice, for any source, we have upper and lower limits for illumination intensities or $I_j^{\min} \leq I_j \leq I_j^{\max}$ for all $j \in \{1, 2, \dots, S\}$. $G_1(\bar{r}; \bar{r}_{s_j})$ is the Green's function which models the propagation of the field from the location of j th source, \bar{r}_{s_j} , to position \bar{r} in the medium. By combining the last two equations, we arrive at:

$$\phi_s(\bar{r}_{d_i}) \simeq \sum_{j=1}^S I_j \int d^3\bar{r} G_2(\bar{r}_{d_i}; \bar{r}) G_1(\bar{r}; \bar{r}_{s_j}) \eta(\bar{r}). \quad (6)$$

In general, G_1 and G_2 are potentially different functions. For instance, in fluorescence tomography G_1 and G_2 model the propagation of excitation and emission fields, respectively. Since the absorption and scattering coefficients, which appear as parameters in the formulation of these Green's functions, are wavelength-dependent, these two Green's functions are slightly different in this problem. Also, in fluorescence tomography, the function $\eta(\bar{r})$ represents the spatial distribution of fluorescent molecules in the medium. To solve this wave propagation problem, we also need boundary conditions. Nonetheless, different boundary conditions only change the form of Green's functions and the main formulation remains untouched. When the volume is discretized to a set of V voxels (i.e., volumetric pixels), and dimensions of voxels are small relative to the geometry of the scanner, we can use the following approximation for η :

$$\eta(\bar{r}) = \sum_{k=1}^V \eta_k \delta(\bar{r} - \bar{r}_{v_k}). \quad (7)$$

Here, it is assumed that all the scattering potential of a voxel is concentrated at the center of the voxel. \bar{r}_{v_k} is the vector that points to the center of the k th voxel which has the total scattering potential of η_k . For this case:

$$\phi_s(\bar{r}_{d_i}) \simeq \sum_{j=1}^S \sum_{k=1}^V I_j G_2(\bar{r}_{d_i}; \bar{r}_{v_k}) G_1(\bar{r}_{v_k}; \bar{r}_{s_j}) \eta_k. \quad (8)$$

In the discrete space, the scattering problem is then formulated by:

$$\bar{\phi}_s = \bar{W}\bar{\eta} + \bar{\varepsilon}, \quad (9)$$

In our notation, $(\bar{\cdot})$ is a matrix, and $(\bar{\cdot})$ and $(\bar{\cdot})^T$ are column and row vectors, respectively. Here, $\bar{\eta} = [\eta_1, \dots, \eta_V]^T$, $\bar{I} = [I_1, \dots, I_S]^T$, $\bar{\phi}_s = [\phi_s(\bar{r}_{d_1}), \dots, \phi_s(\bar{r}_{d_D})]^T$, and $\bar{W}_{D \times V}$ is:

$$\begin{aligned} \bar{W}_{D \times V} &= \bar{G}_2 \cdot \text{diag}(\bar{G}_1 \cdot \bar{I}), \\ \bar{G}_2 &= \begin{bmatrix} G_2^{11} & G_2^{12} & \dots & G_2^{1V} \\ \vdots & . & \ddots & \vdots \\ G_2^{D1} & G_2^{D2} & \dots & G_2^{DV} \end{bmatrix}_{D \times V}, \\ \bar{G}_1 &= \begin{bmatrix} G_1^{11} & G_1^{12} & \dots & G_1^{1S} \\ \vdots & . & \ddots & \vdots \\ G_1^{V1} & G_1^{V2} & \dots & G_1^{VS} \end{bmatrix}_{V \times S}, \\ G_2^{ik} &= G_2(\bar{r}_{d_i}; \bar{r}_{v_k}), \quad G_1^{kj} = G_1(\bar{r}_{v_k}; \bar{r}_{s_j}) \end{aligned} \quad (10)$$

For a given source, detector, and voxel coordinations both $G_1(\bar{r}_{v_k}; \bar{r}_{s_j})$ and $G_2(\bar{r}_{d_i}; \bar{r}_{v_k})$ are constant positive coefficients. $\bar{\eta}$ is the estimation vector, and \bar{I} is the illumination pattern. The last term in Eq. (9), $\bar{\varepsilon} \sim \mathcal{N}(0, \bar{R})$, is the measurement noise which has Gaussian distribution and it is added to make the model realistic.

2.2. State estimation

Prior to scanning, we have an initial estimation of the distribution we aim to reveal. By using that estimation and the forward model of Eq. (9), we can predict the outcome of any measurement. Each time a measurement is made, we can update our estimation by combining the model prediction and result of the noisy measurement. This procedure is well aligned with the function of Kalman filter which combines theoretical predictions with noise contaminated partial observations to compute the updated minimum-variance estimation. Obviously, optimal performance is achieved when measurements are designed to provide maximum information about the distribution at each and every measurement. Therefore, the optimal measurement is the one that minimizes the uncertainty, or variance, of estimation variables. Finding the optimal measurement is the first part of each iteration. Once the measurement is made, we can use Kalman theory to update the state of our estimation, before making the next measurement.

Suppose that $\hat{\eta}^{n|n-1}$ and $\hat{\eta}^{n|n}$ represent our prior and posterior estimations of the vector $\bar{\eta}$ at the time of n th measurement. Uncertainties in the prior and posterior estimations are modeled by covariance matrices: $\bar{P}^{n|n-1} = \text{Cov}(\hat{\eta}^{n|n-1})$ and $\bar{P}^{n|n} = \text{Cov}(\hat{\eta}^{n|n})$, respectively. Following the theory of Kalman filter, the posterior estimation is computed from the prior via the following iterative equation:

$$\hat{\eta}^{n|n} = \hat{\eta}^{n|n-1} + \bar{K}^n \left[\bar{\phi}_s^n - \bar{W}^n \hat{\eta}^{n|n-1} \right]. \quad (11)$$

Here, \bar{K}^n is the Kalman gain which is the weight given to the difference between the prediction and the actual observation. A larger gain places more weight on the latest measurement, while a smaller gain mostly ignores the last measurement and relies instead on the prediction derived

from the previous state. We also need to compute the uncertainty of the updated estimation:

$$\bar{P}^{(n|n)} = \left[\bar{U} - \bar{K}^n \bar{W}^n \right] \bar{P}^{(n|n-1)} \left[\bar{U} - \bar{K}^n \bar{W}^n \right]^T + \bar{K}^n \bar{R} (\bar{K}^n)^T. \quad (12)$$

In this equation, \bar{U} is the identity matrix. The main objective of the design is to find the measurement which leads to a posterior estimate $\hat{\eta}^{n|n}$ that is as certain as possible. We use the trace of the covariance matrix as a measure of estimation uncertainty left in the estimation.

$$Tr[\bar{P}^{n|n}] = Tr[\bar{P}^{n|n-1}] - 2Tr \left[\bar{K}^n \bar{W}^n \bar{P}^{n|n-1} \right] + Tr \left[\bar{K}^n \left\{ \bar{W}^n \bar{P}^{n|n-1} (\bar{W}^n)^T + \bar{R} \right\} (\bar{K}^n)^T \right]. \quad (13)$$

Trace of the covariance matrix is the norm which should be minimized by choosing the best measurement via designing the optimal illumination pattern.

3. Proposed methodology

As shown in Eq. (13), $Tr[\bar{P}^{n|n}]$ is a function of the Kalman gain and illumination pattern embedded in the weight matrix, \bar{W}^n . Therefore, to minimize the trace, we need to find the optimal weight matrix and Kalman gain. To better explain this approach, we first study a scanner with only one detector. After this discussion, we expand the theory to the general case of scanners with multiple detectors.

3.1. Optimal illumination pattern

In the single-detector case, Kalman gain is a column vector and the measurement noise is a scalar $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Therefore, Eq. (13) can be simplified to:

$$Tr[\bar{P}^{n|n}] = Tr[\bar{P}^{n|n-1}] - 2\bar{W}^n \bar{P}^{n|n-1} \bar{K}^n + \left[\bar{W}^n \bar{P}^{n|n-1} (\bar{W}^n)^T + \sigma^2 \right] (\bar{K}^n)^T \bar{K}^n. \quad (14)$$

To minimize this trace, we first assume that \bar{W}^n is constant and we minimize $Tr(\bar{P}^{n|n})$ by choosing the best value for Kalman gain. In order to find the optimal Kalman gain, we need to solve the equation $\nabla_{\bar{K}^n} Tr[\bar{P}^{n|n}] = 0$ ($\nabla_{\bar{a}}(b)$: gradient of b w.r.t. \bar{a}). Solving this equation gives:

$$\bar{K}^{*n} = \left[\bar{W}^n \bar{P}^{n|n-1} (\bar{W}^n)^T + \sigma^2 \right]^{-1} \bar{P}^{n|n-1} (\bar{W}^n)^T. \quad (15)$$

Now, given the optimal Kalman gain, we take the derivative of (14) with respect to \bar{W}^n and set it equal to zero to find the optimal weight matrix. Since our ultimate goal here is finding the optimal illumination pattern, it is better if we reformulate the problem to directly search for the optimal illumination pattern, $\bar{I}^{(n)}$.

$$\begin{aligned} \bar{I}^{*n} &= \underset{\bar{I}^n}{\operatorname{argmin}} \frac{1}{2} (\bar{I}^n)^T \bar{A}^{n|n-1} \bar{I}^n + \bar{b}^T \bar{I}^n, \\ \text{s.t. } I_j^{min} &\leq I_j \leq I_j^{max}, \quad \forall j. \\ \bar{A}^{n|n-1} &= (\bar{K}^n)^T \bar{K}^n \bar{G}^T \bar{P}^{n|n-1} \bar{G}, \\ \bar{b} &= -\bar{G}^T \bar{P}^{n|n-1} \bar{K}^n. \end{aligned} \quad (16)$$

Here, \bar{G} is a $V \times S$ matrix that transforms the source illumination vector \bar{I}^n to the weight matrix $\bar{W}^n = (\bar{G}^n \bar{I}^n)^T$.

$$\bar{G} = \begin{bmatrix} G_2^{11} G_1^{11} & \dots & G_2^{11} G_1^{1S} \\ \vdots & \ddots & \vdots \\ G_2^{1V} G_1^{V1} & \dots & G_2^{1V} G_1^{VS} \end{bmatrix}. \quad (17)$$

The optimization problem of (16) is a Quadratic Programming (QP) problem with inequality constraints, which is convex with a single (global) minimum. Thus, to find the minimum of (14), we can break the optimization into a two-step process. In the first step, we assume \bar{W}^n is constant, and we minimize $Tr[\bar{P}^{n|n}]$ by choosing the best Kalman gain. Next, we adopt the computed gain, and we optimize the cost function with respect to \bar{W}^n . We repeat these two steps till convergence, see Algorithm 1. Once convergence is achieved, we upload the computed optimal illumination pattern on the scanner to make the measurement. The result of the measurement is then used to update our estimation of $\hat{\eta}$ and its covariance matrix.

Algorithm 1 Illumination Pattern Optimization

```

1: initialize  $\bar{P}^{1|0}, \hat{\eta}^{1|0}$  based on prior information;
2: while (certainty < desired level) do
3:   set initial  $\bar{I}^n$ ;
4:   while true do
5:      $\bar{W}^n = \bar{G}^n \bar{I}^n$ ;
6:      $\bar{K}^n = (\bar{W}^n \bar{P}^{n|n-1} (\bar{W}^n)^T + \sigma^2)^{-1} \bar{P}^{n|n-1} (\bar{W}^n)^T$ ;
7:     find  $\bar{I}^n$  using (16) and a QP solver;
8:     check stopping criterion;
9:   end
10:  update  $\hat{\eta}^{n|n}$  using (11);
11:  update  $\bar{P}^{n|n}$  using (12);
12: end
  
```

3.2. Optimal illumination distribution

Source location optimization is another determining factor that plays a role in the performance of the scanner. In section 3.1 we assumed that the number of sources is fixed and only their intensities could change. In that case, the optimal source intensities were referred to as the optimal illumination pattern. In this section, we expand that method by making the number of powered sources also an optimization variable.

We first assume that we have N possible locations for sources to be placed. To limit the number of illuminating sources, it is reasonable to have a minimum of m and a maximum of M illuminating points among the given N locations. This implies that during each measurement, the source vector \bar{I}^n is an N -dimensional column vector with L nonzero entries where $m \leq L \leq M$. Therefore, in this approach, the intensity and location of sources are optimized for a given number of sources and we refer to this as the optimal illumination distribution. To solve this optimization problem, we adopt an integer programming approach. We define an N -dimensional indicator vector \bar{v} with binary elements such that $v_j = 0$ when $I_j = 0$ and $v_j = 1$ when $I_j > 0$. Next, we impose the following linear constraint to the optimization problem:

$$v_j I_j^{min} \leq I_j \leq v_j I_j^{max} \quad (18)$$

This constraint enforces I_j and v_j to be zero simultaneously or $I_j^{min} \leq I_j \leq I_j^{max}$ if $I_j > 0$. To enforce the limits on the number of illuminating sources, we add another constraint:

$$m \leq \sum_j v_j \leq M \quad (19)$$

Therefore, the optimization problem for finding the optimal illumination distribution takes the following form:

$$\begin{aligned} \bar{I}^{*n} &= \operatorname{argmin}_{\bar{I}^n} \frac{1}{2} (\bar{I}^n)^T \bar{A}^n |\bar{I}^n|^{n-1} \bar{I}^n + \bar{b}^T \bar{I}^n, \\ \text{s.t. } v_j I_j^{\min} &\leq I_j \leq v_j I_j^{\max}, \\ m &\leq \sum_j v_j \leq M. \end{aligned} \quad (20)$$

This is a mixed integer quadratic programming problem which is convex and has a unique optimal solution [17].

3.3. Resolution

In general, three factors determine the resolution [18]. The first factor is the transport scattering distance in medium (tissue) which sets a fundamental limit on the spatial resolution. The second factor is the device geometry including the number of sources and detectors. Singular-value analysis has shown that an ideal number of sources-detectors with appropriate distribution, minimizes the linear dependency of the measurements, thereby improving the weight matrix's conditioning and resolution of scanner [6,8,19–21]. The third factor is the prior information, which has a profound impact on resolution [15,22]. For instance, in optical tomography, ϕ_s in Eq. (9) is occasionally measured only from the imaging surface. The dimension of the measured data on the imaging surface is almost always much less than the number of internal nodes in the medium. Therefore, solving the inverse problem for Eq. (9) is ill-posed and hypersensitive to noise. A small signal disturbance may lead to a large reconstruction error which degrades the resolution. Regularization is the common method for handling ill-conditioning of inverse problems [23–27]. In our statistical framework, regularization can be interpreted as an explicit form of incorporating prior information in the formulation of the problem. Adding prior information can induce estimation bias in the same way regularization would do. The statistical reinterpretation of regularization and the introduced bias have not been fully explored yet [28]. However, in a conceptual manner, regularization methods truncate data in directions where corresponding singular values are negligible. Prior information also suggests that we have little uncertainty in some directions and no measurements is required along such directions. This omission in both procedures could result in a bias in the reconstructed image. In other words, the quality of prior information has a profound impact on desired final bias.

Here, we define resolution as the maximum number of voxels that can be revealed at a given certainty when the scanner performs at its best capability. For a fixed number of sources and detectors, the number of discernible measurements is fixed. By increasing the number of voxels that span the space, we increase the number of variables. For a fixed number of measurements, increasing the number of variables reduces the best achievable certainty. A desired certainty is achieved if the number of voxels is less than an upper limit which defines resolution of a scanner.

To measure resolution, we assume that we perform raster scanning which covers all possible measurements. What we measure in the j th measurement, $j \in \{1, \dots, S\}$, is:

$$\begin{aligned} \phi^j &= \bar{W}^j \bar{\eta} + \varepsilon, \\ \bar{W}^j &= (\bar{G} \cdot \bar{I}^j)^T, \\ \bar{I}^j &= [0, 0, \dots, 1, \dots, 0]^T, \varepsilon \sim \mathcal{N}(0, \sigma^2), \end{aligned} \quad (21)$$

where only the j th element of \bar{I}^j is 1. We can combine all these measurements in one equation $\bar{\phi}^B = \bar{W}^B \bar{\eta} + \bar{\varepsilon}^B$ so that $\bar{\phi}^B = [\phi^1, \phi^2, \dots, \phi^S]^T$, $\bar{W}^B = [(\bar{W}^1)^T, (\bar{W}^2)^T, \dots, (\bar{W}^S)^T]^T$, and $\bar{R}^B = \operatorname{diag}[\sigma^2, \sigma^2, \dots, \sigma^2]$. For this case, we have the prior covariance matrix $\bar{P}^{1|0}$. From the

Kalman filter theory, we know that:

$$(\bar{P}^{n|n})^{-1} = (\bar{P}^{n|n-1})^{-1} + \bar{W}^T \bar{R}^{-1} \bar{W}. \quad (22)$$

Therefore, in our case, we have only one update which is:

$$\bar{P}^{1|1} = \left[(\bar{P}^{1|0})^{-1} + (\bar{W}^B)^T (\bar{R}^B)^{-1} \bar{W}^B \right]^{-1}. \quad (23)$$

We can use Eq. (23) to determine the resolution of a scanner for a given prior covariance matrix and noise statistics. Equation (23) shows how our method incorporates both the scanner design and prior information as two resolution enhancing factors. The scanner geometry is embedded in the weight matrix \bar{W} which is optimized based on the method described in section 3.2, and the prior information is revealed in the form of the covariance matrix $\bar{P}^{n|n-1}$.

3.4. State-dependent noise

To this point, we assumed that measurements are contaminated only by Gaussian noise which is state-independent. In reality, the variance of noise in detectors changes as a function of signal amplitude. An example of this effect is the multiplicative noise of photodetectors which implies that noise power increases by the intensity of exposure [29]. In such systems, the additive measurement noise should be modeled by the superposition of state-dependent noise and state-independent Gaussian noise. To add state-dependent noise to the model, we modify the measurement equation as follows:

$$\phi^n = \bar{W}^n(\bar{\eta} + \bar{\varepsilon}_1^n) + \varepsilon_0^n, \quad (24)$$

where $\varepsilon_0 \sim \mathcal{N}(0, \sigma_0^2)$ is the Gaussian and $\bar{\varepsilon}_1$ is the zero-mean state-dependent noise which is linearly coupled to the state vector $\bar{\eta}$ such that:

$$\bar{\varepsilon}_1^n = \sum_i \bar{C}_i \bar{\eta} \mu_i^n. \quad (25)$$

Here, $\mu_i \sim \mathcal{N}(0, \sigma_1^2)$ and \bar{C}_i is a diagonal matrix with only one nonzero element:

$$\bar{C}_1 = c \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{bmatrix}, \bar{C}_2 = c \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{bmatrix}. \quad (26)$$

The coefficient $c \in \mathbb{R}$ is the state-dependent noise gain (SDNG). On trial n , we have a prior estimation $\hat{\eta}^{n|n-1}$, with uncertainty $\bar{P}^{(n|n-1)}$, and we make an observation ϕ^n to update the estimation:

$$\hat{\eta}^{n|n} = \hat{\eta}^{n|n-1} + \bar{K}^n(\bar{W}^n \bar{\eta} + \varepsilon_0^n + \bar{W}^n \sum_i \bar{C}_i \bar{\eta} \mu_i^n - \bar{W}^n \hat{\eta}^{n|n-1}). \quad (27)$$

The covariance of our posterior estimation is:

$$\begin{aligned} \bar{P}^{n|n} &= (\bar{U} - \bar{K}^n \bar{W}^n) \bar{P}^{n|n-1} (\bar{U} - \bar{K}^n \bar{W}^n)^T \\ &+ \bar{K}^n \left[\sigma_0^2 + \sigma_1^2 \sum_i \bar{W}^n \bar{C}_i \bar{\eta} \bar{\eta}^T \bar{C}_i (\bar{W}^n)^T \right] (\bar{K}^n)^T. \end{aligned} \quad (28)$$

This equation states that the uncertainty now depends on $\bar{\eta}$ which is the state-vector. Setting the derivative of the trace of covariance matrix in (28), with respect to $K^{(n)}$, equal to zero results

in:

$$\bar{K}^{*n} = \bar{P}^{n|n-1}(\bar{W}^n)^T \times \left[\bar{W}^n \bar{P}^{n|n-1}(\bar{W}^n)^T + \sigma_0^2 + \sigma_1^2 \sum_i \bar{W}^n \bar{C}_i \bar{\eta} \bar{\eta}^T \bar{C}_i (\bar{W}^n)^T \right]^{-1}. \quad (29)$$

To find the optimal illumination pattern for the system that suffers from state-dependent noise, we take the same approach described in section 3.1. Given the fact that $\bar{W}^n = (\bar{G} \cdot \bar{I}^n)^T$, we can write this optimization problem for the single-detector scanner as:

$$\begin{aligned} \bar{I}^{*n} &= \underset{\bar{I}^n}{\operatorname{argmin}} \quad \frac{1}{2} (\bar{I}^n)^T \bar{A}^{n|n-1} \bar{I}^n + \bar{b}^T \bar{I}^n, \\ \text{s.t.} \quad & I_j^{\min} \leq I_j \leq I_j^{\max}, \quad \forall j, \\ \bar{A}^{n|n-1} &= (\bar{K}^n)^T \bar{K}^n \bar{G}^T (\bar{P}^{n|n-1} + \sigma_1^2 \sum_i \bar{C}_i \bar{\eta} \bar{\eta}^T \bar{C}_i) \bar{G}, \\ \bar{b} &= -\bar{G}^T \bar{P}^{n|n-1} \bar{K}^n. \end{aligned} \quad (30)$$

In practice, we replace the term $\bar{\eta} \bar{\eta}^T$ in Eqs. (29) and (30) with the estimated matrix $\hat{\eta} \hat{\eta}^T$. Notice that due to the addition of state-dependent noise, the optimal Kalman gain, \bar{K}^{*n} , and optimal illumination pattern, \bar{I}^{*n} , depend on the current state of $\hat{\eta}$ which is not the case when the noise is state-independent. For a time-invariant system, in the absence of state-dependent noise, one can compute all optimal patterns in advance if there exists reliable prior information. This method speeds up the data acquisition time since no online processing of the acquired data or updating of the state-vector $\hat{\eta}$ is necessary. This is one of the main advantages of the proposed method which has never been studied before, to the best of our knowledge.

3.5. Multi-detector scanner

In previous sections, we discussed the problem of a single-detector scanner. In this section, we expand the problem to the general case of multi-detector scanners. Once again, we use the trace of covariance matrix, Eq. (12), as a measure of uncertainty to be minimized. Similar to the previous approach, we start the optimization by assuming that the weight matrix, \bar{W}^n , is constant and we minimize $\operatorname{Tr}(\bar{P}^{n|n})$ with respect to Kalman gain.

$$\bar{K}^{*n} = \bar{P}^{n|n-1}(\bar{W}^n)^T \left[\bar{W}^n \bar{P}^{n|n-1}(\bar{W}^n)^T + \bar{R} \right]^{-1}. \quad (31)$$

Finding the optimal illumination pattern in a multi-detector scanner is more complicated and results in the following form:

$$\begin{aligned} \bar{I}^{*n} &= \underset{\bar{I}^n}{\operatorname{argmin}} \quad -2\operatorname{Tr} \left[\bar{K}^n \bar{W}^{(n)} \bar{P}^{n|n-1} \right] + \operatorname{Tr} \left[(\bar{K}^n \bar{W}^n) \bar{P}^{n|n-1} (\bar{K}^n \bar{W}^n)^T \right], \\ \text{s.t.} \quad & I_j^{\min} \leq I_j \leq I_j^{\max}, \quad \forall j. \end{aligned} \quad (32)$$

where:

$$\bar{W}^n = \bar{G}_2 \cdot \operatorname{diag}(\bar{G}_1 \cdot \bar{I}^n). \quad (33)$$

This is not a convex problem, and we do not have a routine procedure to find the global optimum. Nonetheless, since the objective function in Eq. (32) is differentiable with respect to $\bar{I}^{(n)}$, it is reasonable to adopt a variation of the gradient descent method which can also handle the inequality constraint. Fortunately, the method of Projected Gradient Descent (PGD) perfectly tailors to this constrained optimization problem. Starting from an initial point \bar{I}_0 , the gradient descent algorithm iterates the following equation and continues until a stop condition is met:

$$\bar{I}_{q+1} = \bar{I}_q - \alpha_q \nabla_{\bar{I}_q} \operatorname{Tr}(\bar{P}^{n|n-1}). \quad (34)$$

The parameter α_q is the step size, and q is the iteration counter. Stop-condition is met when $\|\bar{I}_{q+1} - \bar{I}_q\| \leq \delta$ where δ is the error threshold. In our simulation experiments we considered

$\delta = 0.01$. Because of the constraint on the optimization problem, a projection step is added to the gradient descent algorithm:

$$\bar{I}_{q+1} = P_r \left[\bar{I}_q - \alpha_q \nabla_{\bar{I}_q} \text{Tr}(\bar{P}^{n|n-1}) \right], \quad (35)$$

where $P_r(\cdot)$ is the projection operator, see algorithm 2).

Algorithm 2 Projected Gradient Descent (PGD)

- 1: Pick an initial vector \bar{I}_0 ;
 - 2: **while** stop condition is not met **do**
 - 3: Pick the descent direction as $-\nabla_{\bar{I}_q} \text{Tr}(\bar{P}^{n|n-1})$;
 - 4: Pick a step size α_q ;
 - 5: Update: $\bar{I}_{q+1} = \bar{I}_q - \alpha_q \nabla_{\bar{I}_q} \text{Tr}(\bar{P}^{n|n-1})$;
 - 6: Projection: $\bar{I}_{q+1} = \min [\max(\bar{I}_{q+1}, \bar{I}_{min}), \bar{I}_{max}]$;
 - 7: **end**
-

Once the solution to problem (32) is found, we use Eq. (33) to calculate the optimal weight matrix, \bar{W}^{*n} . Then, we assume \bar{W}^n is constant to compute the optimal Kalman gain using Eq. (31). We continue till convergence.

4. Results and discussion

Simulations were conducted to demonstrate the performance of algorithms discussed earlier in the article. To better display the concepts, most presented simulations here are in low dimensions; however, without loss of generality, one can expand the same simulations to higher dimensions. In these simulations, we modeled two (2D) or three-dimensional (3D) scanners and we assumed κ is purely imaginary. Imaginary κ means wave propagation is in evanescent mode which is similar to diffusion. Medium was modeled by the diffusion, D , and absorption, μ_a , coefficients. We used the following Green's functions, for 2D and 3D scanners for both G_1 and G_2 [15]:

$$\begin{aligned} G_{2D}(\bar{r}; \bar{r}') &= \sqrt{\frac{2\delta}{\pi|\bar{r} - \bar{r}'|}} \frac{\exp(-\frac{|\bar{r} - \bar{r}'|}{\delta})}{2\pi\mu_a\delta^2}, \\ G_{3D}(\bar{r}; \bar{r}') &= \frac{1}{4\pi\mu_a\delta^2} \frac{\exp(-\frac{|\bar{r} - \bar{r}'|}{\delta})}{|\bar{r} - \bar{r}'|}. \end{aligned} \quad (36)$$

Here, $\delta = \sqrt{D/\mu_a}$. We started our simulations by modeling a 2D single-detector scanner shown in Fig. 3. The scanner has circular geometry with 20 sources distributed uniformly around the circle. The area within the circle was discretized to 31 triangular pixels. We considered this small number of pixels to demonstrate concepts only. In reality, sampling of the medium should be done at much higher rates to achieve the required accuracy. Scattering potentials of these pixels are unknown variables for which we have prior estimation of their mean values and the corresponding covariance matrix. We simulated 20 rounds of optimal measurements/updates. The trace of the covariance matrix was computed after each update and normalized by the initial trace value to compute the Relative Residual Uncertainty (RRU). Curves in Fig. 3 show the evolution of RRU values when the scanner was following the optimal illumination pattern algorithm as well as the conventional raster scanning. This data proves that the optimal illumination approach, compared to raster scanning, expedites the scanning process and generates more accurate images while taking a smaller number of measurements.

This improvement in scanning performance can be justified intuitively as well. Consider a special case where we have no constraint on source intensities while solving the single-detector

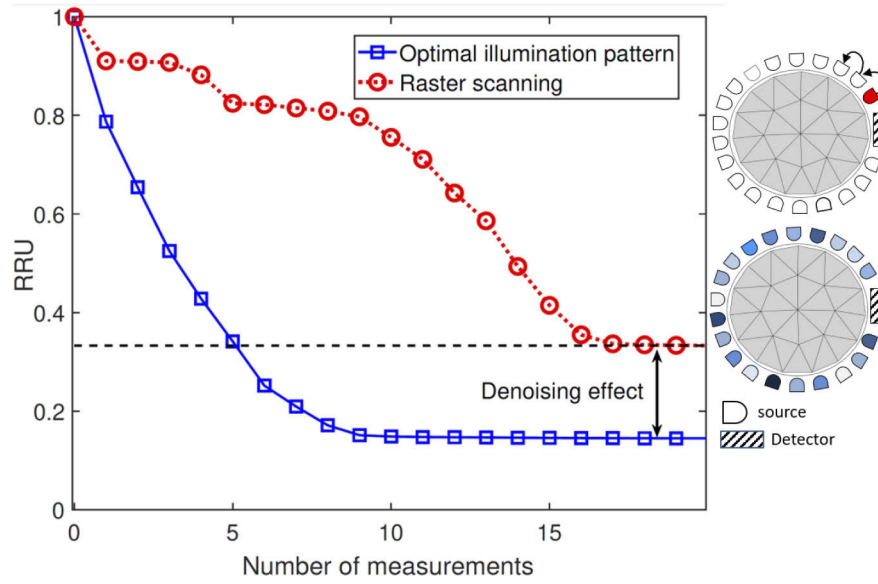


Fig. 3. Structure of a 2D circular single-detector scanner and the comparison between computed RRU values for optimal illumination and raster scanning. Following optimal illumination algorithm, one can reconstruct images with superior certainty while taking a smaller number of measurements.

optimization problem. In this case, the equation $\nabla_{\tilde{W}^{(n)}} \text{Tr}(\tilde{P}^{n|n}) = 0$ has a closed-form solution for the optimal weight matrix:

$$\tilde{W}^{*n} = [(\tilde{K}^n)^T \tilde{K}^n]^{-1} (\tilde{K}^n)^T = (\tilde{K}^n)^\dagger. \quad (37)$$

If we replace this weight matrix in Eq. (15) to compute the Kalman gain, we get $\tilde{P}^{n|n-1} \tilde{K}^n = \lambda \tilde{K}^n$ where λ is a real number. In other words, in each round of measurements if we follow this optimized scanning algorithm, \tilde{K}^n and \tilde{W}^n are both eigenvectors of the prior covariance matrix. This result matches the theory of principal components in which maximum uncertainty occurs along eigenvectors of the covariance matrix. By choosing $\tilde{W} = (\tilde{K}^n)^\dagger$ in each round, we measure the projection of the state vector, $\tilde{\eta}$, along one of its major principal components. As a result, each measurement leads to maximum possible drop in RRU values. In reality, practical constraints on source intensities and scanner geometry make the synthesis of such optimal measurement vectors infeasible. However, when constraints are added to the problem, the proposed method generates illumination patterns that are as close as possible to principle components to help the scanner acquire the most informative projections in each round of measurements. Obviously, this procedure translates to a steeper drop in RRU values compared to raster scanning.

In a single-detector scanner, zero uncertainty and reconstruction error can be achieved in the absence of noise and only if the number of sources is more than or equal to the number of voxels. Obviously, these are not realistic conditions and, as shown in Fig. 3, RRU curves do not cross zero. Nonetheless, intuitively, we expect the final value of RRU, achieved by raster scanning, to be similar to or better than what is ultimately obtained by optimal patterns since any possible illumination pattern is a linear combination of patterns generated in raster scanning. However, as it is shown in Fig. 3, in the presence of measurement noise, optimal illumination algorithm leads to a smaller final RRU value. Note that the covariance matrix update Eq. (12) consists of two parts. The first part is the impact of current measurement and the second part is the effect of the noise covariance matrix. As mentioned before, when there is no constraint on source

intensities, K^n and W^n are along the largest principal component of the state vector. As a result, the term $K^n \bar{R} (K^n)^T$ in Eq. (12) is the noise uncertainty projected on the direction of current measurement. If this noise component is larger than the next largest principal component of the data, the same measurement is repeated to reduce the effect of noise. This is an important fact that is not considered in other scanning methods such as raster scanning. This denoising effect, as displayed in Fig. 3, improves the performance of the optimal illumination algorithm compared to raster scanning both in the scanning speed and the ultimate RRU value.

To better understand this effect, consider a simple 2D single-detector scanner with only two sources and two pixels, Fig. 4. Here, $\bar{\eta}$ is a 2D vector and we have only two principal components. Practical limitations on source intensities set a feasible range for synthesizable measurement vectors. Therefore, in this case, the algorithm designs a measurement vector that is as close as possible to the main principal component. When the noise power in the direction of current measurement is larger than the uncertainty along the second principal component of the state-vector, the optimal measurement vector is positioned along the direction which compromises between the noise component and the second principal component of the covariance matrix to reduce the RRU as much as possible, Fig. 4.

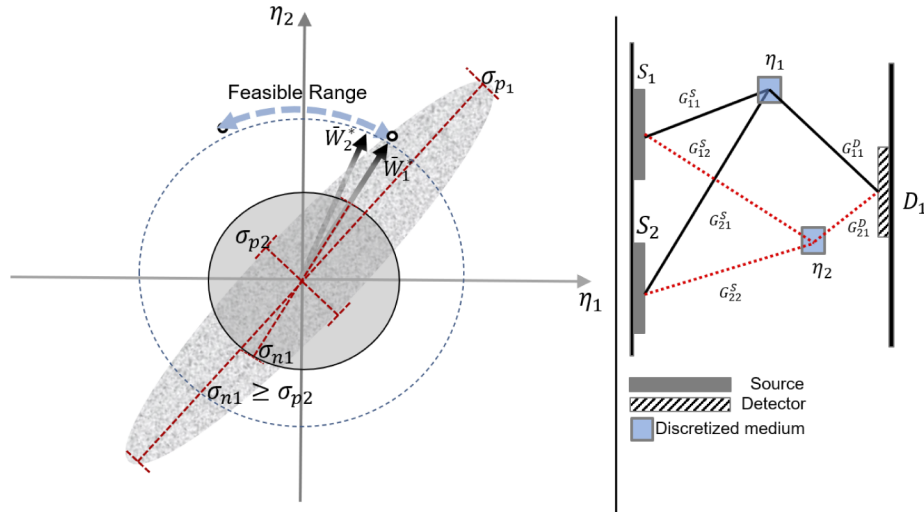


Fig. 4. Effect of noise power on optimal illumination design: (right) structure of a 2D scanner used for the demonstration, (left) the algorithm starts by choosing the first optimal weight vector, \bar{W}_1^* , as close as possible to the first principal component. Since noise power is larger than the uncertainty along the second principal component, the second optimal weight vector, \bar{W}_2^* , chosen by the algorithm is close to the first measurement. In other words, power and distribution of noise suggest that repeating the measurement along the first principal component helps reduce uncertainty more than making the second measurement along the second principal component. This effect leads to superior performance, particularly when noise power is comparable to residual uncertainty along some directions. σ_{p1} and σ_{p2} are the variances along the first and second principal components, respectively. Noise variance, σ_{n1} is uniform in all directions.

In the next test, we included both state-dependent and Gaussian noise in simulations of the same circular 2D scanner. For comparison, we considered four different scenarios in which collected data in all were contaminated by both sources of noise. In simulations, we first reconstructed the image by using optimal illumination algorithm that incorporates or ignores the effect of state-dependent noise. Next, we ran simulations assuming that data is produced via raster scanning and we used the Kalman filter algorithm to update the estimation once incorporating

and then ignoring the effect of state-dependent noise in the reconstruction procedure. Figure 5 displays graphs of final RRU values as a function of SDNG, as described in section 3.4 Eq. (25), for all four different scenarios. Curves in this figure show that the optimal illumination algorithm that incorporates the effect of state-dependent noise outperforms others. Nonetheless, even raster scanning, when the effect of state-dependent noise is taken into account, functions better. Existence of state-dependent noise is a realistic assumption for most detection systems, and as graphs show, considering the effect in scanning and image reconstruction can lead to noticeable improvements.

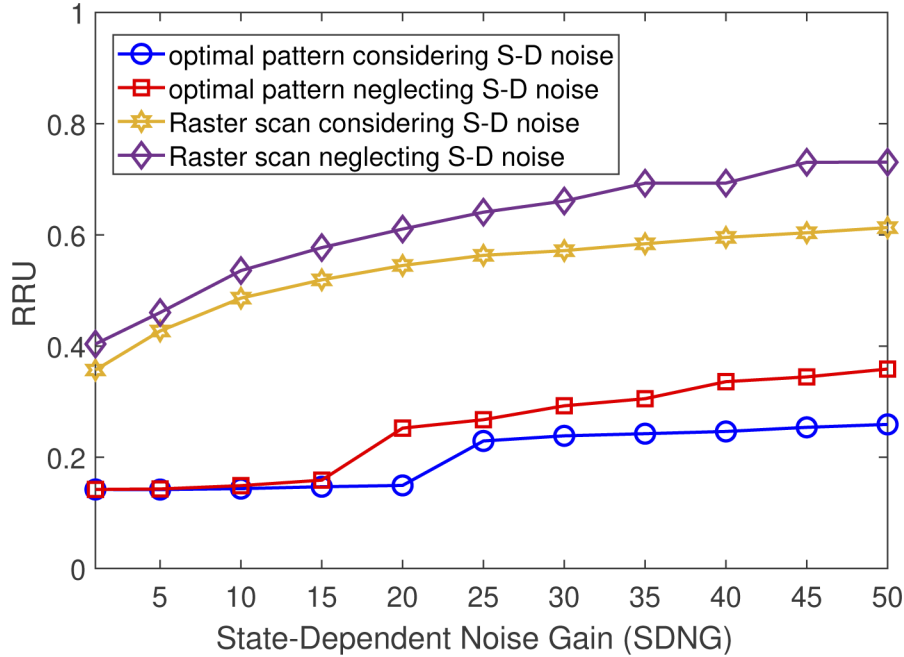


Fig. 5. Effect of state-dependent noise on final RRU values plotted as a function of SDNG for four different data acquisition scenarios. Data shows optimal illumination protocol, when state-dependent noise is included in the model, outperforms other algorithms.

According to our discussion on the resolution, both prior information and data acquisition capability of the scanner (determined by the geometry of the scanner and the relative position of sources, detectors, and voxels) have a significant impact on achievable resolution. To demonstrate these effects, we simulated an 8cm^3 cubical scanner with the geometry shown in Fig. 6(a). Two spherical objects with Gaussian spatial distribution ($\sigma = 0.05$) and 0.5cm apart center-to-center were positioned in the middle of the phantom. As stated in section 3.3, to measure resolution, we assume that we perform raster scanning which covers the maximum capability of the scanner. The number of sources multiplied by the number of detectors ($S \times D$) was increased gradually and RRU values were computed for three different prior information and several different voxel numbers. To investigate the effect of prior information, three different initial covariance matrices (poor, good, and excellent) were generated for each number of voxels. For this purpose, we considered the three parameters p_{middle} , σ_r , and σ_m for the initial estimations. p_{middle} is the probability of objects being in the middle one-eighth of the cube, σ_r is the variance of the spherical distribution, and σ_m is the variance of distribution magnitude. Each of these parameters can get three different values for poor, good, and excellent prior information, according to Table 1. Next, for each category, 1000 random spatial distributions were generated and then covariance matrix and state vector were initialized. To quantify the quality of prior information, we define Normalized Initial

Trace (NIT) as the trace of the initial covariance matrix relative to the worst initial trace value (poor). Three separate curve sets in Fig. 6(b) show that for given prior information, the number of voxels that can be reconstructed at a specific certainty level is directly proportional to the number of sources multiplied by the number of detectors ($S \times D$). Meanwhile, if we start with more certain prior information, we can achieve better resolution for a fixed $S \times D$ value. For instance, based on this data, maximum number of voxels to be reconstructed at 90% certainty is 512 for $S \times D = 114$ with excellent prior information. On the other hand, with poor prior information, we need about $S \times D = 210$ to achieve the same resolution.

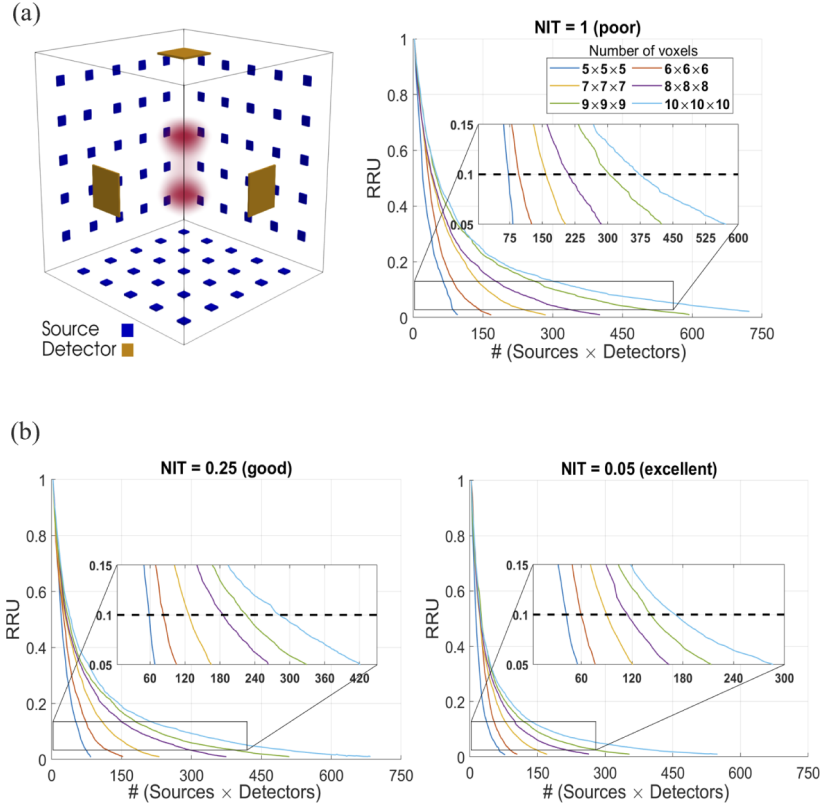


Fig. 6. Analysis of main factors impacting resolution in the statistical framework: (a) geometry of the cubical scanner, (b) relationship between the quality of prior information, number of sources-detectors, and achievable resolution.

Table 1. Characteristics of three different prior estimations

Prior information quality	p_{middle}	σ_r	σ_m	NIT
Poor	0.2	0.1	4.1	1
Good	0.5	0.08	2.1	0.25
Excellent	0.8	0.06	0.9	0.05

To better display the effect of prior information on resolution, we performed the simulation with 64 sources and 12 detectors mounted on opposing walls of the cubic scanner while performing raster scanning. The medium inside the scanner was spanned to $12 \times 12 \times 12$ voxels. We simulated the tomography process once by starting with no prior estimation and then we repeated

with much less uncertainty in our prior information close to the actual state. Figure 7 shows the result of these reconstructions. As displayed, having proper prior information significantly enhances the resolution of reconstructed images for a fixed $S \times D$ and two spheres are readily identified as distinct objects.

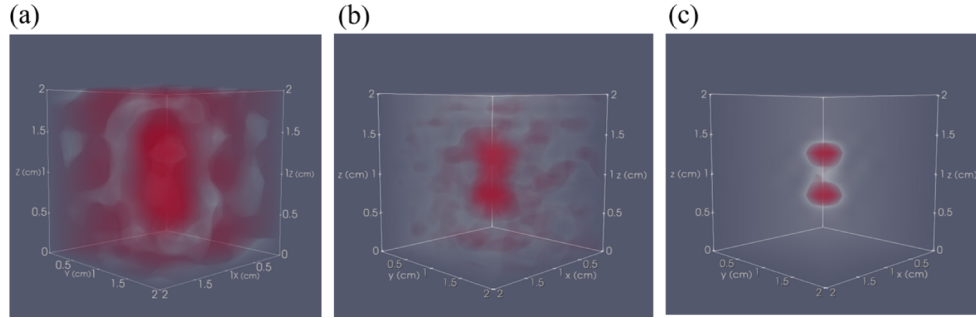


Fig. 7. Reconstruction of two spherical objects: (a) no prior information, (b) poor prior information about the location/distribution of targets, (c) adequate prior information, quality and resolution are significantly improved.

In the next part, a simple experiment was designed to show the effect of source locations on the amount of information acquired from each measurement using an optimal illumination pattern. A 2D circular scanner with a single source and single detector was designed. We assumed that the location of detector is fixed at 0° , and the medium under the test was discretized to 10 pixels for illustration the concept. We let pixels with more uncertainties be located in a specific region of the scanning area (dark pixels in Fig. 8(a)) to highlight the effect of source location on the scanner's capability. Two different schemes were examined. First, a single source could freely move all around the scanner while illuminating the optimal power. RRU value was calculated at each angle to see which angular location reduces uncertainty the most. Four rounds of measurement were performed while in each round, the estimation was updated using the best source location. In the second test, we assumed that the single source could only be placed in a predefined location in each round of measurements; i.e., location of the single source is at 0° in the first measurement and it rotates 90° counter-clockwise in each next round. Results in Fig. 8(a) show that RRU value significantly changes as the source moves around the scanner. It is also shown that there is always an optimum location in every measurement that can gather the most amount of information and reduce uncertainty as much as possible, Fig. 8(b).

To assess the performance of the illumination distribution optimization method, discussed in section 3.2, We ran simulations using the same 2D scanner. Here, another freely rotating source was added to the scanner geometry. The two light sources could freely change their locations at 72 different angles around the circle. RRU values were computed for every two combinations of 72 source locations, and a color map was generated with each source location on one axis. Gurobi optimizer [30] was then used to solve the optimization problem (20) and select the best

pair of source locations among the $\binom{72}{2}$ existing options. Results illustrated in Fig. 8(c) show that the proposed method is perfectly capable of finding optimal source locations that minimize RRU the most.

Ultimately, to evaluate the performance of a multi-detector scanner, we conducted simulations using an inhomogeneous digital mouse phantom [31]. Two spherical fluorescence targets, with an approximate radius of 0.8mm and Gaussian distributions, were placed inside the mouse brain. Optical properties of the tissue were adjusted to the values given in Table 2. The mouse head was

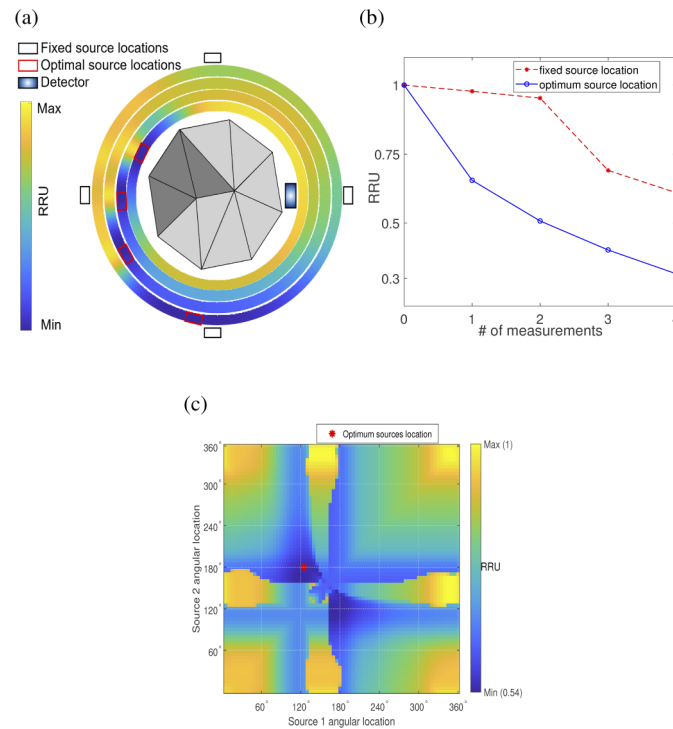


Fig. 8. Effect of source location on acquired information: (a) First, a source rotates around the scanner and RRU is computed at each location. Rings show color-coded values of RRU for each angular position. Red rectangles show locations where RRU is minimum. Next, RRU was computed for fixed predefined locations (black). (b) Quantitative analysis of experiment of panel (a). Selecting optimal location results in a considerable decrease in RRU, (c) Color map of RRU computed for every two combinations of 72 source locations. The red marker indicates the optimal location.

tessellated into 3375 voxels. 36 sources and 30 detectors were placed on the surface of the head as shown in Fig. 9(a). Panel (b) in this figure shows the cross-section of the phantom and the distribution of targets at the depth of $z = 12\text{mm}$. Optimal illumination pattern was obtained for each measurement using PGD algorithm with an adaptive step size and the estimation was updated after each observation. The adaptive method starts from 0.002 and chooses a different step size at each iteration based on the difference between the trace of covariance matrix in the last two consecutive iterations. Raster scanning was also performed for comparison. Figure 9(c) shows the evolution of reconstructed cross-sections at the same depth for both optimal illumination pattern and raster scanning for different number of measurements. Results show clearly that the optimal illumination algorithm outperforms raster scanning. The two targets could be identified after making 18 measurements using optimal patterns, while raster scanning required 36 measurements to reconstruct images of fluorescence objects with comparable clarity. Although optimization of illumination pattern for each measurement requires large computation power, advances in technologies such as super computers significantly reduce computation time. Therefore, using our online reconstruction method, reducing the number of measurements decreases reconstruction time as well. For instance, if the number of measurements is reduced by half compared to raster scanning, the reconstruction time is also cut in half.

In raster scanning, sources are turned on one after another without any policy about how more information can be acquired. It is shown that sometimes random scanning may lead to

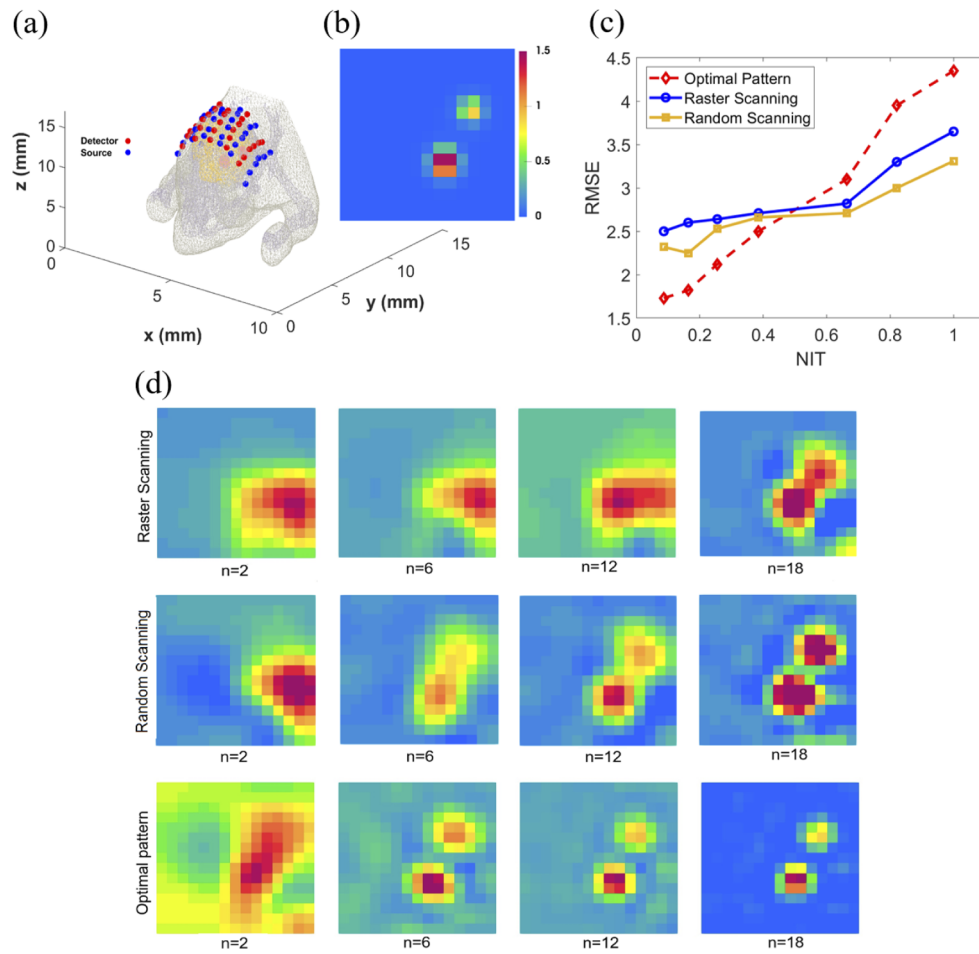


Fig. 9. (a) Digital mouse phantom simulation setup. (b) Cross-section of digital mouse phantom at $z = 12\text{mm}$. (c) effect of prior information on reconstruction error in three different scanning protocols. (d) Cross-sections of fluorescence targets reconstructed via raster scanning (top row), random scanning (middle row), and optimal pattern method (bottom row) shown for different number of measurements. The proposed method is able to reconstruct while taking less number of measurements.

Table 2. Optical properties of digital mouse phantom

Tissue Type	Brain	Skull	Skin
$\mu'_s(\text{cm}^{-1})$	12.5	10.0	8.0
$\mu_a(\text{cm}^{-1})$	0.178	0.101	0.159

better performance (compared to raster scanning). The simplest practical random scanning pattern would be randomly turning on single sources which at least ensures the exploration of perpendicular directions. In our experiment, shown in Fig. 9(d), random scanning outperforms simple raster scanning. However, the proposed optimal scanning protocol shows superior performance. It is worth noting that because of the non-convex nature of the multi-detector problem and inefficiency of current methods in finding global optimum, there may be other patterns such as random patterns that have better partial performance.

Finally, we performed a quantitative analysis on the reconstructed images for the three examined scanning protocols. Clearly, quality of prior information has a profound impact on the performance of the proposed optimum scanning protocol. This effect can be seen in the final reconstruction error. To examine this effect, several prior estimations with different qualities were considered, and the reconstruction error was calculated in each case. To quantify the quality of prior information, we calculated Normalized Initial Trace (NIT) as the trace of the initial covariance matrix relative to the worst initial trace value. As it is shown in Fig. 9(c), unreliable prior information not only may lead to improper illumination patterns, but also results in larger bias in the final reconstructed image compared to other two scanning protocols.

5. Conclusion

Scanning the sample under test is the first essential step in tomography and potentially the most time-consuming part of the process. During scanning, the sample receives multiple dosages of exposure which could be harmful in some forms of tomography such as X-ray imaging or when radioactive materials are exploited. The article concentrates on formulating a systematic approach to take advantage of the available computational power and optimize the scanning process via minimizing the necessary number of measurements to achieve acceptable image quality. Developed algorithms also open the door for optimizing the geometry of the scanner or the location of sources/detectors to obtain superior performance. We incorporated determining factors such as state-independent and -dependent measurement noise and prior information about the sample in our formulations. At this stage, the method is developed for static samples which are not evolving during the scanning process or speed of changes are negligible compared to the optimal scanning speed. Nonetheless, the proposed method has the potential to expand to dynamically evolving samples where scanner keeps changing its configuration and illumination patterns based on online computations to maximize the amount of acquired information in each step. The proposed method can be utilized to optimize the data acquisition process in any experimental study including different tomography modalities such as electrical impedance tomography, thermo-acoustic tomography, etc. Whenever a model of the system is obtained based on our knowledge of the underlying physics, this method can improve the efficiency of measurements and the overall performance of the system. In our simulations, we only investigated the reconstruction of reduced scattering coefficient from the data. To obtain simultaneous reconstructions of scattering and absorption coefficients, intensity measurements alone are proven to be insufficient and data from frequency or time domain are needed.

Currently, the main limitation of our method is that it requires high computational power to search for the optimal patterns. For instance, increasing the number of voxels results in a large covariance matrix that makes it extremely challenging to find the optimum pattern with the present computation power. The good news is with growing advances in optimization methods and computational power, such as cloud computing and super computers, finding the optimal solution to such nonlinear problems will be more trivial in the near future.

Funding. National Science Foundation (1454300, 1830145); Army Research Office (W911NF181032).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. S. Sabir, C. Kim, S. Cho, D. Heo, K. H. Kim, J. C. Ye, and S. Cho, "Sampling scheme optimization for diffuse optical tomography based on data and image space rankings," *J. Biomed. Opt.* **21**(10), 106004 (2016).
2. S. Belanger, M. Abran, X. Intes, C. Casanova, and F. Lesage, "Real-time diffuse optical tomography based on structured illumination," *J. Biomed. Opt.* **15**(1), 016006 (2010).
3. J. Ripoll, "Hybrid Fourier-real space method for diffuse optical tomography," *Opt. Lett.* **35**(5), 688–690 (2010).
4. N. Ducros, C. Dandrea, G. Valentini, T. Rudge, S. Arridge, and A. Bassi, "Full-wavelet approach for fluorescence diffuse optical tomography with structured illumination," *Opt. Lett.* **35**(21), 3676–3678 (2010).
5. A. Joshi, W. Bangerth, K. Hwang, J. Rasmussen, and E. M. Sevick-Muraca, "Plane-wave fluorescence tomography with adaptive finite elements," *Opt. Lett.* **31**(2), 193–195 (2006).
6. J. P. Culver, V. Ntziachristos, M. J. Holboke, and A. G. Yodh, "Optimization of optode arrangements for diffuse optical tomography: a singular-value analysis," *Opt. Lett.* **26**(10), 701–703 (2001).
7. H. Xu, H. Dehghani, B. W. Pogue, R. Springett, K. D. Paulsen, and J. F. Dunn, "Near-infrared imaging in the small animal brain: optimization of fiber positions," *J. Biomed. Opt.* **8**(1), 102–110 (2003).
8. E. E. Graves, J. P. Culver, J. Ripoll, R. Weissleder, and V. Ntziachristos, "Singular-value analysis and optimization of experimental parameters in fluorescence molecular tomography," *J. Opt. Soc. Am. A* **21**(2), 231–241 (2004).
9. L. Hao, G. Li, and L. Lin, "Optimization of measurement arrangements for magnetic detection electrical impedance tomography," *IEEE Trans. Biomed. Eng.* **61**(2), 444–452 (2014).
10. J. Dutta, S. Ahn, A. A. Joshi, and R. M. Leahy, "Illumination pattern optimization for fluorescence tomography: theory and simulation studies," *Phys. Med. Biol.* **55**(10), 2961–2982 (2010).
11. M. Bergounioux, É. Bretin, and Y. Privat, "How to position sensors in thermo-acoustic tomography," *Inverse Probl.* **35**(7), 074003 (2019).
12. Y. Liu, W. Ren, and H. Ammari, "Robust reconstruction of fluorescence molecular tomography with an optimized illumination pattern," *Inverse Probl. & Imaging* **14**(3), 535–568 (2020).
13. M. Althobaiti, H. Vavadi, and Q. Zhu, "Diffuse optical tomography reconstruction method using ultrasound images as prior for regularization matrix," *J. Biomed. Opt.* **22**(2), 026002 (2017).
14. G. Zhang, H. Pu, W. He, F. Liu, J. Luo, and J. Bai, "Bayesian framework based direct reconstruction of fluorescence parametric images," *IEEE Trans. Med. Imaging* **34**(6), 1378–1391 (2015).
15. B. Zhang, S. Liu, X. Cao, F. Liu, X. Wang, J. Luo, B. Shan, and J. Bai, "Fluorescence tomography reconstruction with simultaneous positron emission tomography priors," *IEEE Trans. Multimedia* **15**(5), 1031–1038 (2013).
16. A. Ale, V. Ermolayev, E. Herzog, C. Cohrs, M. H. De Angelis, and V. Ntziachristos, "FMT-XCT: in vivo animal studies with hybrid fluorescence molecular tomography–X-ray computed tomography," *Nat. Methods* **9**(6), 615–620 (2012).
17. D. Bienstock, "Computational study of a family of mixed-integer quadratic programming problems," *Math. programming* **74**(2), 121–140 (1996).
18. F. Leblond, H. Dehghani, D. Kepshire, and B. W. Pogue, "Early-photon fluorescence tomography: spatial resolution improvements and noise stability considerations," *J. Opt. Soc. Am. A* **26**(6), 1444–1457 (2009).
19. T. Lasser and V. Ntziachristos, "Optimization of 360 degree projection fluorescence molecular tomography," *Med. Image Anal.* **11**(4), 389–399 (2007).
20. F. Leblond, K. M. Tichauer, and B. W. Pogue, "Singular value decomposition metrics show limitations of detector design in diffuse fluorescence tomography," *Biomed. Opt. Express* **1**(5), 1514–1531 (2010).
21. C.-W. Chen and Y. Chen, "Optimization of design parameters for fluorescence laminar optical tomography," *J. Innovative Opt. Health Sci.* **04**(03), 309–323 (2011).
22. S. C. Davis, H. Dehghani, J. Wang, S. Jiang, B. W. Pogue, and K. D. Paulsen, "Image-guided diffuse optical fluorescence tomography implemented with Laplacian-type regularization," *Opt. Express* **15**(7), 4066–4082 (2007).
23. A. N. Tikhonov and V. Y. Arsenin, "Solutions of ill-posed problems," *SIAM Rev.* **21**(2), 266–267 (1979).
24. J. Baritoux, K. Hassler, and M. Unser, "An efficient numerical method for general l_p regularization in fluorescence molecular tomography," *IEEE Trans. Med. Imaging* **29**(4), 1075–1087 (2010).
25. L. Zhao, H. Yang, W. Cong, G. Wang, and X. Intes, " L_p regularization for early gate fluorescence molecular tomography," *Opt. Lett.* **39**(14), 4156–4159 (2014).
26. J. Shi, F. Liu, G. Zhang, J. Luo, and J. Bai, "Enhanced spatial resolution in fluorescence molecular tomography using restarted L_1 -regularized nonlinear conjugate gradient algorithm," *J. Biomed. Opt.* **19**(4), 046018 (2014).
27. D. Hyde, E. L. Miller, D. H. Brooks, and V. Ntziachristos, "Data specific spatially varying regularization for multimodal fluorescence molecular tomography," *IEEE Trans. Med. Imaging* **29**(2), 365–374 (2010).
28. D. Calvetti and E. Somersalo, "Inverse problems: from regularization to Bayesian inference," *WIREs Comp. Stat.* **10**(3), e1427 (2018).
29. A. Ullah, W. Chen, M. A. Khan, and H. Sun, "A new variational approach for multiplicative noise and blur removal," *PLoS One* **12**(1), e0161787 (2017).
30. L. Gurobi Optimization, "Gurobi optimizer reference manual," (2021).

31. B. Dogdas, D. Stout, A. F. Chatzioannou, and R. M. Leahy, "Digimouse: a 3D whole body mouse atlas from CT and cryosection data," *Phys. Med. Biol.* **52**(3), 577–587 (2007).